DNA barcodes democratize genetics

07/03/2012
Ashley Yeager

With a quick scan of one or two genes, scientists, and students, can ID an animal, plant or fungi they've found in the field or in their backyard. But, as Ashley Yeager reports, not all researchers see the scientific value of the simplified genetic fingerprinting technique.

Kate Stoeckle's dad, Mark, talked about work a lot, so, over sushi, she innocently asked him if scientists could use a simple genetic test he spoke of to ID the fish in front of them. Yes, said Mark Stoeckle, a physician and adjunct professor at Rockefeller University. He had been an advocate of the genetic fingerprinting technique, called DNA barcoding, since 2003. Now, it was 2007 and sequencing technology had developed enough to allow him to help his daughter and her friend, both high schoolers at the time, to figure out what fish were served and sold at their favorite New York City restaurants and supermarkets.

DNA barcodes of four species, two similarly looking butterflies and two types of owl. Credit: University of Guelph.

Stoeckle had explained to his daughter that DNA barcoding was a system that sequences the same part of the same gene of every creature on Earth and stores the information in the Barcode of Life Database. Scientists could then match their sequenced gene region with similar scans to identify the species of their sample. Through their experiment, the Stoeckle team found that restaurants faked tilapia as more expensive white tuna and labeled the endangered Acadian redfish as red snapper (1). The study sparked others like it (2), which demonstrated not only rampant fish fraud, but also how DNA barcoding is rapidly becoming like a "toaster, a technology everyone can use," Stoeckle said.

For regulating consumer markets, DNA barcoding might be as good an application of the system as there is, according to Daniel Rubinoff, an entomologist at the University of Hawai'i at Manoa. In these cases, there are "just a relative handful of well-known species and their DNA relationships are well-established, so barcodes can offer a quick shortcut," he said. But not all animals' and plants' molecular relationships are as well understood. So, when it comes to identifying species with additional genetic complexity, like Karner blue butterflies (3), comparing creatures' life history based on only one gene can't give amateur or professional scientists the information they need to accurately identify species (4).

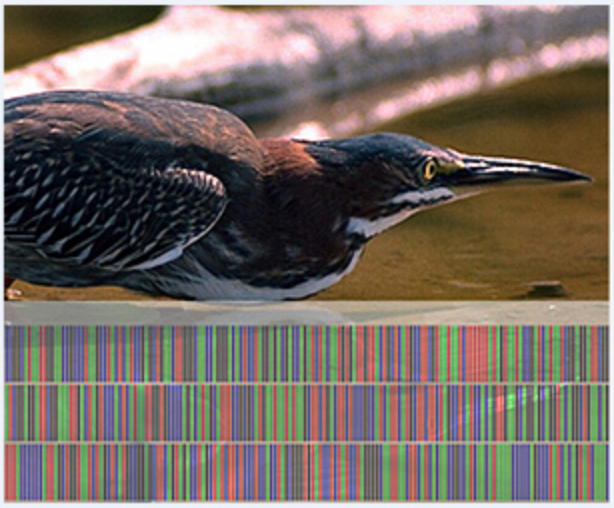
"The point in science is accuracy. When it coincides with simplicity, that's great. But in the case of barcoding, it's giving people who don't know better a false sense of knowledge," said Rubinoff, adding that most critics of DNA barcoding are scientists who study the history of life on Earth.

DNA barcoding isn't necessarily for this group of scientists, though, according to biologist Paul Hebert of the University of Guelph in Ontario, Canada. He is considered the father of DNA barcoding. In 2003, he published a description of the method, which promised a fast, efficient way to sequence a gene or two of a specimen and identify its species (5). No longer would biologists and naturalists need to consult an expert taxonomist each time they stumbled across a whole animal or plant, or even pieces of one, that they did not recognize, he said. Instead, scientists, or any interested citizen, could collect a sample and either send it off for sampling at a lab, which costs about $25, or sequence the barcode genes themselves, for about $5, and get a basic idea of what they had found. The system would be similar to the way a check-out clerk scans the barcodes comprised of stripes and rings up a grocery bill. But instead of barcodes linking to price, these would be barcodes linking to life, said Hebert.

To build this database quickly and cheaply, barcode scientists sequence one gene in animals and two genes in plants. In animals, they target a 648 base pair region of the cytochrome *c* oxidase subunit 1 (COI) gene of mitochondria. COI barcoding does not work in most species of plants, however, because the rate of evolution of the gene is slower, making it more difficult to distinguish plant species based on variation in this gene region. As a result, scientists use two chloroplast genes, rbcL and matK, as the plants' barcodes (6).

The process to sequence DNA barcodes in both animals and plants is similar, despite the use of different genes as the universal barcodes. First, scientists, or students with access to sequencing machines, have to isolate the DNA from their sample. The barcoders then amplify the target gene regions in the extracted DNA using PCR. "PCR is absolutely critical to DNA barcoding right now," said Hebert. "It delves into the total DNA extract of the sample and winnows out all the other sequences except the one or two genes we need for a barcode."

The PCR primers used are in conserved sequences that flank the *rbc*L and COI barcode regions, which can have some variation between individuals in a species. Because the regions have accumulated some sequence differences over evolutionary time, it is impossible to identify a



This heron is shown with its "DNA barcode," which provides biologists with information about its relationship to other animals. Credit: Biodiversity Institute of Ontario

universal primer set for *rbc*L or COI that will work across all taxonomic groups of plants and animals, respectively. The experimental protocol for DNA barcoding, which is publicly available, lists several standard PCR primers, ensuring students and scientists from around the world can amplify the correct DNA region and then have it sequenced to obtain the DNA barcode and identify their sample from it.

So far, researchers have used the available primers to generate barcodes for 114,365 animal species, 39,730 plant species and 2,366 fungi and other life forms. These numbers, however, only begin to "scratch the surface" of the five to 10 million eukaryotic species that scientists estimate live on Earth and in its oceans, said Hebert. Because there are so many more to identify, he and his collaborators have launched a worldwide effort — through the International Barcode of Life initiative — to build the barcode library to 500,000 species by 2015. Another, more distant, goal is also to build a hand-held sequencer so that field researchers could scan a sample and possible identify a new species based on similar matches in the database of DNA barcodes.

It is this use of DNA barcoding that concerns Rubinoff and others the most. What doesn't make much sense about DNA barcoding, he said, is to assume that there is "any single gene, and any one part of that single gene which could accurately reflect the complexity of the process of speciation across all of animal life." Evolutionary biologists understand that the "process of speciation is a gradual one, like pulling taffy apart. Not a cleaver coming down, suddenly and abruptly on one entity, suddenly making it two," said Rubinoff. He added that, as sequencing technology develops, the idea that "we'd use a few hundred base pairs of one gene to understand the relationships between all of life when we have so much more information easily available is going to seem absurd. Like looking at a panorama through a pinhole."

Hebert, Stoeckle and other advocates of DNA barcoders are well aware of their critics. But barcoding advocates are less interested in the technical disadvantages and more interested in the democratizing and educational uses of DNA barcoding, which are still developing, they argue. And next-generation sequencing, said Hebert, will only speed efforts to build the barcode library, while making the entire process cheaper and more efficient. "We won't be just tearing legs off bugs, or scraping tissue from fish or leaves," he said. "We'll all be able to take water and amplify the barcode genes from a mess of organisms, and we'll really see what is, or was, around in this world."

References:

1. Stoeckle, K., and Strauss, L. High school students track down fish fraud. Pacific Fishing. September 2008.

2. Hubert, N. et. al. Barcoding Freshwater Fishes. (2008). PLoS One. 3:6, e2490

3. Gombert, Z. et. al. Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. (2006) Molecular Ecology. 15, 1759–1768.

4. Song, H. et. al. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. (2008). PNAS. 105:36, 13486-13491.

5. Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. deWaard. Biological identifications through DNA barcodes. (2003) Proc. Roy. Soc. Lond. Ser. B. 270, 313-321.

6. CBOL Plant Working Group. A DNA barcode for land plants. (2009). PNAS 106 (31): 12794–12797.